



City Research Online

City St George's, University of London

Citation: Christophel, T. B., Allefeld, C., Endisch, C. & Haynes, J-D. (2018). View-Independent Working Memory Representations of Artificial Shapes in Prefrontal and Posterior Regions of the Human Brain. *Cerebral Cortex*, 28(6), pp. 2146-2161. doi: 10.1093/cercor/bhx119

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/22845/>

Link to published version: <https://doi.org/10.1093/cercor/bhx119>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

View-independent working memory representations of artificial shapes in prefrontal and posterior regions of the human brain

Thomas B. Christophel^{a,b,c,*}, Carsten Allefeld^{a,b,c}, Christian Endisch^{a,b} & John-Dylan

Haynes^{a,b,c,d,e,f}

a Bernstein Center for Computational Neuroscience, Charité Universitätsmedizin, Berlin, 10115, Germany

b Berlin Center for Advanced Neuroimaging, Charité Universitätsmedizin, Berlin, 10117, Germany

c Clinic for Neurology, Charité Universitätsmedizin, Berlin, 10117, Germany

d Berlin School of Mind and Brain, Humboldt Universität, Berlin, 10099, Germany

e Cluster of Excellence NeuroCure, Charité Universitätsmedizin, Berlin, 10117, Germany

f Department of Psychology, Humboldt Universität zu Berlin, Berlin, 10099, Germany

* Corresponding Author: tbchristophel@gmail.com, +49 173 545 9701

Running Title: View-Independent Working Memory Storage

Keywords: Working Memory, Visual Short-Term Memory, fMRI, Multivariate Pattern Analyses, Invariance

Acknowledgments: This work was funded by the Bernstein Computational Neuroscience Program of the German Federal Ministry of Education and Research BMBF Grant 01GQ0411, the Excellence Initiative of the German Federal Ministry of Education and Research DFG Grants GSC86/1-2009, KFO247 and HA 5336/1-1.

Author Contributions: TBC & JDH designed the study with help from CE. TBC & CE acquired the data. The data was analyzed by TBC & CA. The paper was written by TBC, CA & JDH.

Abstract

Traditional views of visual working memory postulate that memorized contents are stored in dorsolateral prefrontal cortex using an adaptive and flexible code. In contrast, recent studies proposed that contents are maintained by posterior brain areas using codes akin to perceptual representations. An important question is whether this reflects a difference in the level of abstraction between posterior and prefrontal representations. Here we investigated whether neural representations of visual working memory contents are view-independent, as indicated by rotation-invariance. Using fMRI and multivariate pattern analyses, we show that when subjects memorize complex shapes, both posterior and frontal brain regions maintain the memorized contents using a rotation-invariant code. Importantly, we found the representations in frontal cortex to be localized to the frontal eye fields rather than dorsolateral prefrontal cortices. Thus, our results give evidence for the view-independent storage of complex shapes in distributed representations across posterior and frontal brain regions.

Introduction

The human brain is capable of temporarily storing a limited amount of visual information in an active and flexible workspace typically referred to as visual working memory (Baddeley 1986, 1992; Luck 2008). The contents of working memory can be manipulated within this workspace and are believed to serve as input for higher order processes (Logie 2003; Hyun and Luck 2007). For a long time it was believed that visual working memory relies on a centralized store in dorsolateral prefrontal cortex (DLPFC), mainly based on evidence from electrophysiological studies in non-human primates (Courtney et al. 2007; Funahashi 2006; Goldman-Rakic 1995; Petrides 1995). In contrast, recent work in human neuroimaging using multivariate pattern analyses has raised evidence for multiple separate stores in visual, temporal and parietal brain areas retaining visual information held in working memory (Harrison and Tong 2009; Serences et al. 2009; Christophel et al. 2012; Riggall and Postle 2012; Emrich et al. 2013; Christophel and Haynes 2014), while spatial, categorical and quantitative representations were also found in frontal areas (Jerde et al. 2012; Spitzer and Blankenburg 2012; Lee et al. 2013). The debate over the localization of working memory storage is currently ongoing (Albers et al. 2013; Lee et al. 2013; Jacob and Nieder 2014; Lara and Wallis 2014; Matsushima and Tanaka 2014; Mendoza-Halliday et al. 2014; Christophel et al. 2015, 2017; Bettencourt and Xu 2016).

However, while research has focused on the *localization* of the working memory stores, it has remained unclear which *format* of storage is used. In visual areas, for example, information might be retained in a strictly retinotopic format mimicking the early visual responses during perception and retaining the specific viewing conditions during encoding. Alternatively, the memory representation of visual contents could exhibit some degree of *invariance* meaning that

it generalizes across viewing conditions similar to perceptual representations of objects in lateral occipital cortex (Grill-Spector et al. 2001; Kourtzi and Kanwisher 2001; Kourtzi et al. 2003; see also Williams et al. 2008; Cichy et al. 2011).

Here, we intended to shed light on both format and localization of visual working memory contents. In particular, we asked whether representations of memorized contents generalized across different rotational views and whether this depended on whether the rotation information was irrelevant or critically necessary for the task (See Figure 1). To ensure that our results were not confounded by verbal encoding strategies, we refrained from using real-life objects as stimuli. Instead, we used complex artificially generated shapes that are hard to describe verbally. Notably, neural representations of *simple* shapes (like circles, triangles and squares) have been found during working memory in monkey dorsolateral prefrontal cortex (DLPFC; Meyer et al. 2011; Meyers et al. 2012). Delay-period information about similarly *complex* shapes, however, was only investigated and found in inferotemporal areas (Miyashita and Chang 1988). The stimuli we used here allowed us to parametrically vary shape similarity between memorized samples and test stimuli and thereby adjust task difficulty (See Figure 2).

We particularly asked whether it is possible to identify invariant representations of these complex shapes in regions beyond early visual cortices. Two groups of subjects memorized shapes shown in random in-plane rotations while being positioned in an MRI scanner. By using random rotations rather than a set of fixed rotations we ensured that subjects had to re-encode the samples in each trial discouraging long-term memory based strategies (i.e. ‘object A in rotation 3’). Please note that as a disadvantage our use of random rotations prevents us from applying simple and direct tests of rotation-specific information. The confirmation of rotation-specific encoding schemes was only a secondary interest of the current study, as representations that

encode individual rotations have been identified before (e.g. Harrison and Tong 2009; Christophel et al. 2015).

The two groups performed two different versions of a delayed shape similarity task. One group performed a *rotation-invariant task* where the rotation of the memorized shape was irrelevant whereas subjects from a second group performed the *rotation-specific task* and had to remember its rotation in addition to its form (see Figure 1 and its caption). The tasks were performed by separate groups to avoid a carryover of encoding strategies between the two groups. To identify whether the brain retained these shapes using a rotation-invariant code, we used multivariate pattern analysis (MVPA, Haxby et al. 2001; Haynes and Rees 2005a, 2005b; Kamitani and Tong 2005; Haynes and Rees 2006; Harrison and Tong 2009; Serences et al. 2009; Christophel et al. 2012; Riggall and Postle 2012; Lee et al. 2013) to test whether the BOLD-fMRI patterns encode shape-specific information in a form that generalizes across different rotational states.

Materials and Methods

Participants

44 healthy right-handed subjects (25 female; mean age: 23.9, SEM \pm 0.53) with normal or corrected-to-normal vision were recruited for the current study. Subjects gave informed consent and the study was approved by the local ethics committee. The subjects were split into two equally-sized groups either performing a rotation-invariant (group 1: 14 female; mean age: 23.4, SEM \pm 0.65) or a rotation-specific (group 2: 11 female; mean age: 24.4, SEM \pm 0.83) version of the task.

Artificial shape stimuli

In both experiments, participants were asked to memorize artificially generated shapes. Stimuli were presented via back-projection using Cogent 2000 (<http://www.vislab.ucl.ac.uk/cogent.php>). We used decagons, ten-sided shapes shown in white on a black circle (4° visual angle in diameter; Figure 2A) on a grey background (not shown in the Figures). These artificial shape stimuli were chosen to discourage memorizing the stimuli using verbal descriptors (i.e. for real-life objects) while at the same time presenting stimuli of challenging complexity. These stimuli varied in the distance of the 10 corners to the center of the shape. To generate these shapes, we assigned one out of 10 evenly-spaced numbers (from 0.28 to 1) randomly to each corner. All points were 36° radial angle apart from each other relative to the center. We chose decagons over shapes with more or less sides, as ten-sided shapes were sufficiently hard to memorize. To control for slight variations in center-of-mass and surface area, all stimuli were then recentered and resized to a surface area of 3 square degrees of visual angle. Please see Figure 2A for a visual representation of this correction.

For each subject in the first experiment, we randomly generated a unique set of six decagons (sample stimuli, Figure 1, *Design*), selected from a larger set of random shapes to generate a diverse set. Trying to avoid multiple close-to-identical shapes, we estimated similarity for all 15 pairs of six shapes regardless of rotation, correlating the original (not recentered) corner-to-center distances of potential shapes for all pairs of possible rotations in steps of 36° (rotation-invariant correlation, Figure 2B). For a given set of six shapes, we thus considered 150 correlation values between pairs of ten corner-to-center distances (10 per pair), which inevitably included some spurious correlations leading us to choose a lenient criterion of Fisher z-transformed $r < .55$ (corresponding to $r < .5005$) for the inclusion of shapes. Even this lenient

criterion limited the number of ten-sided shapes within a set to six, as the unsupervised random generation algorithm was unable to find larger sets within a reasonable runtime. The same criterion was used to assure that different rotational versions of the same shape were largely dissimilar. The same 22 sets were used for the 22 subjects in the second experiment to allow for a better comparison between the two groups.

Delayed shape similarity tasks

In each trial, subjects memorized one out of the six possible shapes. To separate memory-related signals from neural correlates of mere stimulus processing or visual persistence we used a retro-cue method (Figure 1; Sperling 1960; Harrison and Tong 2009; Christophel et al. 2012). At the beginning of the trial, two decagons shown after each other in rapid succession followed by a retro-cue (either a '1' or '2') that indicated which of the two was to be remembered (see Figure 1 for timings). The cue was surrounded by a star-shaped 20-sided mask that was introduced to suppress perceptual representations of the samples. The cued sample was to be held in memory for the following delay of 9.5 seconds.

To probe for rotation-invariance, the memorized shape was shown in a randomly chosen in-plane rotation. Specifically, we used a different rotation in each trial to avoid a limited set of predetermined rotational states, which subjects might recognize. As a result, potential non-invariant retinotopic representations of a given memorized shape would not induce a common activation pattern across trials and the neural signals for the six shapes would remain impossible to differentiate for multi-voxel pattern analyses. This was further enforced by controlling the memorized samples for size and center-of-mass and by keeping correlations between stimuli and amongst rotational versions of the same stimulus below a common criterion (see above). In

comparison, invariant representations that can generalize across viewing conditions should leave a distinguishable neural trace across stimuli memorized in various rotations.

To serve as a non rotation-invariant control, the non-remembered shape was always shown in the same rotation. This allowed us to directly compare potential rotation-invariant working memory representations to perceptual representations without necessitating invariant processing during perception. We ensured during piloting that subjects would not be aware of this manipulation and subjects confirmed this in the post-scan debriefing.

After the delay, two test stimuli were shown left and right of the center of the screen (Figure 1, right). Subjects were asked to identify which of the two test stimuli was more similar to the memorized sample. None of the stimuli was identical to the remembered stimulus and both target and foil were generated randomly. The target stimulus was chosen to have a positive correlation and the foil was uncorrelated (see below).

Importantly, the two subject groups performed two slight variations of the task: In the first group, subjects performed a rotation-invariant version of the task (Figure 1, top right). In this group, the target stimulus rotation was chosen randomly and was similar to the memorized shape when rotated optimally (rotation-invariant z-transformed $r > .6$, see Figure 1, for examples). The similarity of the target stimulus was adjusted for each subject during training to approximate threshold performance (see below). In contrast, the foil stimulus was uncorrelated regardless of its rotation (rotation-invariant z-transformed $r < 0.2$). Thus, the stimulus rotation during encoding was irrelevant for this task and a rotation-invariant representation would be sufficient to perform the task.

In the second group, subjects had to memorize the shape of the cued sample in the specific rotation shown to perform in the task (Figure 1). For this purpose, the target stimulus was always shown in a rotation corresponding to the memorized shape (rotation-specific z-transformed $r > .6$, see Figure 1, for examples). As in experiment 1, the exact similarity of the target stimulus was adjusted for each subject during training to approximate threshold performance. Furthermore, the foil's correlation to the memorized sample was only kept below criterion when computed for the rotational states shown (rotation-specific z-transformed $r < .2$, see Figure 1, for examples). Thus, the foil's shape could be more similar to the memorized shape but shown in a different rotation. Finally, we included a number of catch trials (25%) where the foil was identical to the memorized stimulus but shown in a random rotation. The rotations of these catch foils were sampled from the full 360° space such that subjects were sometimes asked to identify deviations of a few degrees (average deviation 90° , uniform distribution). This approach did prevent us from using the rotation-specific correlation approach to quantify similarity as the center-to-corner angles did not align between the memorized sample and the foil. These catch trials allowed us to assess behaviorally whether subjects retained the rotation information throughout the delay.

In both groups, subjects performed four runs of 60 trials each in the fMRI scanner, 10 per memorized stimulus. Prior to the experiment in the fMRI scanner, subjects received one hour of training to familiarize them with the difficult task. For this training, we shortened the delay period to 2.5 seconds and included feedback. Sample shapes were generated randomly for each trial to avoid the training of a stimulus-set specific strategy. Furthermore we used the training to titrate the difficulty of the target stimulus using three levels of difficulty (both experiments, range of 0.6 to 1.35 in z-transformed correlations) and an intermediate level of difficulty was

used for the scanning session that approximates 75% performance based on linear interpolation. On average, this difficulty was 1.02 (SEM ± 0.03) in the rotation-specific task and 1.16 (SEM ± 0.02) for the rotation-invariant task. Please note that, due to different methods of calculating rotation-invariant and rotation-specific similarity (see Figure 2) these values and the resulting indicators of behavioral performance are not easily compared across groups. Subjects received one run of training without feedback directly before the scanner experiment.

For comparison with recently observed content-specific working memory signals in the frontal eye fields (see Jerde et al. 2012), subjects in the second group additionally performed a very simple eye movement task used to localize the frontal eye fields (Ferrier 1875). Subjects were instructed to fixate unless one out of four symbols told them to make either horizontal or vertical eye movements for two seconds. In particular, we used the letter ‘h’ and the ‘=’ symbol to instruct horizontal eye movements while ‘v’ and ‘||’ signified vertical eye movements. By avoiding visually-guided saccades we insured that responses during this task were specific to eye movements and not to the visual stimulation used to instruct them. Subjects were asked to move their eyes back and forth as far as they comfortably could (*ad lib.*) and with as many repetitions as possible during the 2 second cue presentation. The inter-cue interval lasted either two, four or six seconds with an average of four seconds. Each subject performed four runs of 20 trials each, five for each cue. Activity during this task was used to identify eye-movement related activation at the putative location of the frontal eye fields.

Data acquisition

For the main experiment, subjects performed four runs of 60 trials each in the fMRI scanner (Siemens 3T Tim Trio). During each run, we recorded 547 three-dimensional images of

BOLD data using an EPI sequence (TR: 2 s, TE: 30 ms, voxel size 3.3x3x3 mm, slice gap: 0.33 mm, descending order, FOV: 192 mm). The 33 slices were aligned approximately parallel to the anterior and posterior commissures and covered the whole neocortex. We timed the experiments in such a way that each trial onset coincided with the acquisition onsets of a functional image (stimulus-locked acquisition) thereby reducing variance in the temporal signature of the task. In addition, we acquired anatomical data (T1-weighted MPRAGE: 192 sagittal slices, TR: 1900 ms, TE: 2.52 ms, flip angle: 9°, FOV: 256 mm).

fMRI preprocessing

Functional imaging data was analyzed using SPM8, cvMANOVA (Allefeld and Haynes 2014) and The Decoding Toolbox (Hebart et al. 2015). After conversion to NIFTI format, the functional data were motion corrected and the anatomical image was coregistered to the first image of the BOLD time series. No normalization into a standard space was performed at this time and we applied no Gaussian smoothing to the data prior to performing the multivariate analyses to preserve the fine-scaled spatial structure of the fMRI data. Using a similar reasoning, we abstained from using slice-time correction during preprocessing to retain the temporal precision of the stimulus-locked time-series (see 'Data Acquisition').

Analyses of multivariate pattern distinctness using cvMANOVA

To identify representations of the remembered shapes and to further test these representations for invariance, we used a recently developed technique for multi-voxel pattern analysis technique, cross-validated MANOVA (cvMANOVA, Allefeld and Haynes 2014). The approach is similar to recently used classifier-based ‘decoding’ analyses (Haynes and Rees

2005a, 2005b; Kamitani and Tong 2005; Haynes and Rees 2006; Soon et al. 2008; Sterzer et al. 2008; Harrison and Tong 2009; Serences et al. 2009; Cichy et al. 2011; Christophel et al. 2012; Riggall and Postle 2012; Lee et al. 2013; Christophel et al. 2015) but has several advantages: (1) It quantifies patterned activation differences in a continuous manner rather than counting binary ‘true’ or ‘false’ classifications; (2) its result is an interpretable measure of multivariate effect size (explained variance) which can be expected to be more comparable across studies; and (3) it makes the full flexibility of the regressor-based general linear model (GLM) available to multivariate analyses (Friston, Holmes, et al. 1994).

cvMANOVA uses an adaptation of the well-known framework of multivariate analysis of variance (Timm 2002) to quantify differences in the patterned hemodynamic responses for different neural events. BOLD time series are fitted with a multivariate general linear model (MGLM) using a set of regressors of the same kind as those used in a univariate GLM. An effect of interest is specified in the form of a contrast vector or matrix, and the amount of multivariate variance is calculated that can be explained by this effect, as a ratio to the multivariate error variance. Since the naive computation strongly overestimates the true amount of explained variance (a form of the problem of overfitting), cross-validation is used to obtain an unbiased estimate, which is called estimated pattern distinctness \hat{D} . Specifically, cvMANOVA uses leave-one-run-out cross-validation, effectively comparing the pattern difference estimates averaged across all but one run with the pattern difference estimate obtained from the left-out run by computing their covariance; in this view, \hat{D} is a measure of the consistency of patterned responses across runs.

In detail, the MGLM

$$Y = XB + \varepsilon$$

is fitted to the measured time series Y of length n from a set of p voxels, so that Y is an $n \times p$ matrix. The design matrix X models the expected shape of the BOLD signal as a linear superposition of q regressors (an $n \times q$ -matrix). Accordingly, B is a $q \times p$ matrix which describes the influence of the q different regressors in generating the signal in the p different voxels. An effect of interest is specified in the form of a contrast matrix C with q rows and an arbitrary number of columns. The contribution of this effect to the model parameters is then given by $B_\Delta = CC^-B$, where $^-$ denotes the pseudoinverse. The quantity we are interested in is

$$D = \text{trace} \left(\frac{1}{n} B_\Delta' X' X B_\Delta \Sigma^{-1} \right),$$

where Σ is the variance-covariance matrix of the error ε , and $'$ is the matrix transpose. The pattern distinctness D quantifies the amount of multivariate signal variance explained by the effect, in relation to the multivariate error variance. To estimate D , the MGLM is fitted to data from each of m runs separately, providing estimated regression parameters \hat{B} , and estimated errors $\hat{\varepsilon}$ (residuals). Then, leave-one-run-out cross-validation is implemented as follows. The estimate of D in the l th fold is

$$\hat{D}_l = \text{trace}(H_l E_l^{-1})$$

where

$$H_l = \sum_{k \neq l} \{ \hat{B}_\Delta' \}_k \{ X' X \hat{B}_\Delta \}_l \quad \text{and} \quad E_l = \sum_{k \neq l} \{ \hat{\varepsilon}' \hat{\varepsilon} \}_k,$$

with $k=1 \dots m$. Braces with subscript indicate that the design matrix and estimated parameters are to be taken from the respective run. It can be shown that the mean of the fold-

wise estimates \hat{D}_l is an almost-unbiased estimator of D , which can be converted into a perfectly unbiased estimator \hat{D} via an additional factor depending on properties of the experimental design. In the rest of the paper, for simplicity we use the symbol D when referring to the estimated pattern distinctness. For more details of cvMANOVA, please refer to Allefeld and Haynes (2014); an implementation for MATLAB is available online.

Under the null hypothesis of no patterned activation differences (no explained multivariate variance) D has a true value of 0, and estimated values of D vary randomly around 0. Negative values of D are an expression of this random variation. This corresponds to the random variation of estimated accuracies around the chance level (for two classes: 50 %) in the case of classifier-based ‘decoding’. cvMANOVA can be either combined with a searchlight approach (Kriegeskorte et al. 2006) investigating local patterns in spherical portions of the data (here, with a radius of 4 voxels; i.e. in ‘searchlights’ of maximally 251 voxels, not limited to the cortical surface but excluding cerebrospinal fluid) across each position of the brain to obtain a statistical parametric map of D -values, or applied to regions of interest.

In this paper we use cvMANOVA to answer two questions: Do the fMRI data contain information about the *identity* of the memorized shape? And do they contain information about the *rotation* a given shape is presented at? The analyses described in the following were implemented identically for each subject and for both experiments.

Shape identity information: To probe the data for information about the identity of the presented shape, each set of trials corresponding to one of the 6 different shapes was modeled separately. To additionally investigate the time-course of information and to avoid the assumptions of the canonical hemodynamic response function (HRF), time points within the

delay period were modeled separately using a set of 5 FIR regressors (boxcar functions of 2 s length each such that a period of 10 s was covered; this includes the time of the cue display). We therefore had a 6×5 factorial design with factors “shape identity” and “time point”, leading to 30 regressors in total (for each of the 4 runs). Different rotations of the same shape in the ten different trials (per shape and run) were not modeled separately.

Because we were interested in the effect of shape identity, we used a contrast matrix which (with respect to a single time point) has the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix},$$

where the rows correspond to the 6 shape identity regressors and the columns to the five partial contrasts together defining the effect of shape identity. The true pattern distinctness corresponding to this contrast is 0 if the activation patterns are identical for the first and second shape, the second and the third shape, etc., i.e. if all the shapes have the same corresponding activation pattern. A deviation from this leads to an above-zero value of D.

Since there were 5 different regressors for each shape identity corresponding to the 5 time points, this contrast matrix had to be padded with zeros. For the first time point this has the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix},$$

where the rows correspond to the 30 regressors per run, with FIR regressors for each shape grouped together. For each of the 5 resulting contrasts (one per time point) we estimated the amount of multivariate variance explained by the encoded effect relative to the multivariate error variance, resulting in the measure of pattern distinctness D . This procedure was performed separately for each of the 5 time points during the delay period and for all possible searchlights across the brain resulting in 5 brain-shaped maps of D , one for each time-point.

These maps were normalized into MNI space using normalization parameters acquired during unified segmentation (Ashburner and Friston 2005) of the coregistered anatomical images and smoothed using a Gaussian kernel with 3 mm FWHM to reduce residual between-subject variation in the functional anatomy. To be probed for significant deviations from zero, these normalized and smoothed maps of D were then averaged across time and entered into a second-level one-sample t-test. This resulted in statistical maps representing values of t and p , which were thresholded using a two-step cluster-correction approach (Friston, Worsley, et al. 1994; Nichols and Hayasaka 2003). All statistical tests are performed on data that was averaged across the whole delay-period and all bar plot and brain render displays correspond to averaged data.

In this analysis focusing on shape identity we did not separately model the different rotational states of each shape. This way, the multivariate variance induced by differences in rotation could not be accounted for by the MGLM regressors, but contributed to the error variance. Since the pattern distinctness D is the ratio between explained and error variance, an effect of shape identity therefore not only indicates the presence of shape identity information, but also that the amount of rotation-induced variance is not sufficiently high as to prevent significance. Moreover, a different set of 10 rotations were used in each run, and since D can also be interpreted as a measure of pattern consistency across runs, this indicates that the patterns

corresponding to different rotational states of the same shape are not sufficiently different as to prevent the identification of a common underlying pattern associated with the shape only. Both of this implies that a significant result for shape identity information can also be read as an indicator that the patterns representing shape identity are retained in an invariant format across rotational states.

Shape rotation information: To further investigate whether the areas identified in the first analyses do not only contain shape identity information, but store shapes in a rotation-invariant way, we also probed the recorded fMRI data for rotation-specific information. An invariant memory store should have no information about the rotation of the memorized shape, whereas a retinotopic store would necessarily contain *rotation* information, but should show no representation of shape *identity* when collapsed across rotations.

For this analysis we used a different set of MGLM regressors, additionally modeling rotational state. Since rotations were randomly selected for each trial separately and never repeated across runs, we used 3 evenly spaced bins of rotation angles (e.g. for rotations from 0° to 120° , from 120° to 240° and from 240° to 360° , respectively). The boundaries of these rotation bins were chosen separately for each shape, such that the distribution of trials across the 3 bins and the 4 runs was as uniform as possible. We therefore had a $6 \times 3 \times 5$ factorial design with the factors “shape identity”, “shape rotation”, and “time point”, leading to 90 regressors in total (for each of the 4 runs). To test whether the patterned activation differed for different rotations of the same shape, we estimated the pattern distinctness D using a contrast matrix that encodes differences between rotation bins, within each shape. This analysis can be understood as a series of six separate analyses estimating the main effect of rotation for each separately with the resulting pattern distinctness values being averaged afterwards. The resulting maps of D were

processed and tested as in the previous analysis. For one subject in the second group we found no boundaries that resulted in at least one trial per bin for all shapes and rotations, and this subject was discarded from this analysis.

Frontal eye field region-of-interest analysis

The second group also performed a frontal eye field localizer. Eye movement related activity was modeled using regressors based on the canonical HRF. The resulting parameter estimates were tested against baseline on the group level using a one-sample t-test after normalization and smoothing using a kernel of 7 mm FWHM. To yield a conservative localization of the frontal eye fields we used voxel-level correction for multiple comparisons (Friston et al. 1991) and clusters smaller than 20 voxels were removed. We then extracted the frontal portions of these clusters and separated them according to hemisphere resulting in regions-of-interest (ROIs) for the left and right frontal eye fields. After transformation into single-subject space (inverse normalization) to preserve the single-subject resolution of the activity patterns, these ROIs were then probed in both groups for invariant information about the memorized shape using the cvMANOVA approach. These analyses proceeded in exactly the same way as the searchlight analysis, only that voxels were selected based on the identified ROI instead of distance from a center searchlight voxel. The resulting pattern distances were tested on group level against zero using one-sample t-tests.

Multivariate pattern classification using the decoding toolbox

To compare results from the pattern distinctness analyses to a more conventional analyses approach, we used multivariate decoding analyses (Cox and Savoy 2003; Haynes and Rees 2005a; Kamitani and Tong 2005; Kriegeskorte et al. 2006) combined with a time-resolved searchlight approach (Kriegeskorte et al. 2006) to probe the data for rotation-invariant and

rotation-specific representations. For this, we used a publicly available toolbox that was developed in-house (Hebart et al. 2015). More specifically, we used the rotation-bin model described above to identify (a) regions that have representations of shape-identity that generalize across different rotations and conversely (b) regions that represent a shapes rotational state. As a first step, for both analyses, we computed a univariate general linear model to estimate beta estimates for all 360 regressors (6 shapes by 3 rotations by 4 runs by 5 time-points).

Shape identity cross-classification: We estimated rotation-invariant local pattern information regarding the memorized shape identity in small spherical (three-dimensional) clusters with a radius of 4 voxels (i.e. in ‘searchlights’ of maximally 251 voxels) at each position in the brain (not limited to the cortical surface but excluding cerebrospinal fluid). We used parameter estimates from one out three rotation bins for each shape train a linear support vector pattern classification algorithm (LIBSVM, regularization parameter $C=1$) to distinguish between a pair of memorized shapes. This training was performed using data from three out of four runs. The classifier was then applied on the parameter estimates for the remaining, left-out run and the two remaining rotations bins for the respective shapes. Thus, overall classifier performance was based on a series of pair-wise (2-way) trainings of all possible pairs of shapes and rotation bins and tests of the respective classifiers on the remaining rotation bins. Chance level was 50%. The resulting accuracies were then averaged across all classification steps. Each time leaving out a different run, this procedure was repeated four times (4-fold cross validation). The resulting accuracies were averaged across the cross-validation folds and noted in a dedicated brain map at the position of the center voxel of the searchlight. This was repeated for each center position across the brain and for each of the five time-points, yielding five maps of decoding accuracies

for that subject. The accuracy maps for each subject were then normalized, smoothed, averaged and tested as in the previous cvMANOVA analyses.

Shape rotation classification: Rotation-specific local pattern information was calculated within the different beta estimates for a given shape identity. Using data from three out of four runs, we trained classifiers to distinguish between data from two out of three rotation bins and tested these classifiers on data from the remaining run. This procedure was repeated for all shapes, all possible pairs of two rotation bins, all cross-validation folds, all searchlight center voxels, all time-points and the resulting accuracy maps were averaged, post-processed and tested as before.

Results

Behavioral results

The two groups performed two variants of a difficult shape similarity discrimination task where the retention of the rotational view of the shape was either irrelevant (rotation-invariant task, group 1) or necessary (rotation-specific task, group 2) for the task (see Figure 1 and its caption). During the training stage before scanning, the similarity of the target stimulus to the memorized shape was varied in order to set a difficulty level that would lead to around threshold performance. The similarity was adjusted during scanning when subjects performed below threshold. In the scanner, the target was correctly identified in 80.92% (SEM \pm 1.78 %) of the trials in the rotation-invariant task and in 74.13% (SEM \pm 1.59 %) in the rotation-specific task (see Figure 3, for single subject data). In 9.17 % and 10.83 % of the trials (in the invariant and

specific task, respectively), subjects responded too late (after two seconds) or did not respond at all. During a number of catch trials in the rotation-specific task, the foil was identical to the memorized sample in shape but was randomly rotated (see Methods). For these catch trials which included trials with close-to-identical-rotations, accuracy was 65.6 % (SEM \pm 1.5 %) and thus lower (paired t-test, $T_{(21)} = 6.37$, $p < 0.001$) than in standard trials (77.8 %; SEM \pm 1.66 %), demonstrating that subjects used the rotation information for the task, albeit with imperfect precision. Given that catch trials were identical to standard trials all the way up to the target presentation and that our period of interest was the preceding delay-period, our analyses collapsed across all trials.

Content-specific brain activity patterns during the invariant task

First, we wanted to know which brain regions have content-specific representations of the memorized shapes. We assumed that, within an area constituting a visual working memory store, different shapes held in memory would lead to differential neural activity patterns (MVPA, Haxby et al. 2001; Haynes and Rees 2005a, 2005b; Kamitani and Tong 2005; Haynes and Rees 2006; Harrison and Tong 2009; Serences et al. 2009; Christophel et al. 2012; Riggall and Postle 2012; Lee et al. 2013), potentially reflecting a sampling of neural population codes (Fujita et al. 1992; Tanaka 1993) by fMRI voxels (Kriegeskorte et al. 2010). Specifically, we tested whether local spatial patterns of brain activity differed between the different experimental conditions (shapes). For this we used a recently introduced difference measure called *pattern distinctness D* (Allefeld and Haynes 2014). *D* quantifies the information carried by brain signals with respect to the experimental conditions, in this case the shapes (see Methods; for details). Technically, this reflects the multivariate variance explained by a given contrast of conditions in relation to the

residual multivariate variance. Values of D significantly larger than 0 indicate the presence of reliably different patterns and therefore the representation of content-specific information in the respective area.

In the rotation-invariant task, collapsing across rotational views, we found that in multiple regions different memorized shapes evoked significantly different fMRI-patterns as indicated by significantly positive values of D (one-sample t-test, $p < 0.001$, cluster-corrected $p_{(FWE)} < 0.05$, Figure 4A). The regions included occipito-temporal, posterior parietal and frontal brain areas. In occipital cortex, the information was localized on the lateral surface extending into both visual and inferotemporal cortex. The peaks in occipital cortex ($XYZ_{(MNI)}$; [-44 -70 -4] & [46 -82 10], $T_{(21)} = 5.73$ & 4.71) are consistent with the localization of the lateral occipital complex (LOC; Malach et al. 1995; see also Kanwisher et al. 1997; Grill-Spector et al. 1999). Pattern distinctness D in posterior parietal cortex was most reliable in areas close to the intraparietal sulcus ($XYZ_{(MNI)}$; [-22 -72 50] & [34 -50 48], $T_{(21)} = 5.16$ & 5.79), consistent with previous studies investigating working memory of pattern stimuli (Christophel et al. 2012; Christophel and Haynes 2014) and extended into more anterior regions of the parietal cortex. Finally, we found regions in frontal cortex localized in premotor cortices ($XYZ_{(MNI)}$; [-54 8 34] & [52 8 32], $T_{(21)} = 3.75$ & 4.43). Importantly, a similar analysis probing the same data for information about the shown but not memorized shape revealed significantly less information in all three regions (paired t-test, LOC: $T_{(21)} = 3.81$ & 2.14; PPC: $T_{(21)} = 3.29$ & 4.75; FEF: $T_{(21)} = 3.48$ & 2.91, $p < 0.05$, see Figure 5A).

We also tested shape representations outside of these areas using a more sensitive and explorative region-of-interest approach. For this, we used two locations in left and right calcarine

sulcus that were shown to represent mentally rotated stimuli in prior work (Christophel et al. 2015; XYZ_(MNI); [-1 -82 10] [10 -88 10]). Furthermore, we defined two anatomical regions-of-interest in left and right DLPFC (XYZ_(MNI); [-44 52 21] [44 52 21]). Neither of these regions showed significantly larger delay-period information for memorized as compared to not-memorized shapes (see Figure 5A). In the absence of such a differential effect, we assume that the information present in the calcarine sulcus for both memorized and not-memorized items is indicative of a perceptual rather than a working memory representation. Not-memorized items were always shown in the same rotation to serve as a non rotation-invariant control (see Methods) which might result in higher pattern distinctness values in right calcarine sulcus.

For confirmatory purposes we also performed a more conventional cross-classification analyses which has been used previously to study invariance (Cichy et al. 2011). For this purpose, we grouped trials of a given shape with similar rotations into three bins (see Methods; for details). We trained support vector machine classifiers to distinguish between different shapes, but using only data from one of these three rotation bins. Then, we introduced data from the remaining rotation bin into these classifier models to predict the shape identity, essentially asking whether the difference between a pair of shapes in one type of rotation shared a common difference in patterned brain activity with the same pair of shapes in different rotations (see Methods; for details). Using this analyses, we find significant above chance classification in all regions reported above (LOC: $T_{(21)} = 3.87$ & 4.96 ; PPC: $T_{(21)} = 4.57$ & 4.5 ; FEF: $T_{(21)} = 2.95$ & 2.52 , $p < 0.05$, see Figure 5B).

Rotation-specific information versus rotation-invariance

In a brain area that stores the stimulus in a rotation-invariant format, there should be reliable information about the identity of the shape, but no reliable differences between different rotational states of the same shape. In contrast, in a region storing stimuli in a strictly retinotopic format without rotation invariance, identical shapes shown in different rotations would be represented by very different patterns of activity. We thus tested, whether any regions exhibit rotation-specific information by testing the pattern distinctness D for the different rotational angles. The previous analysis that collapsed across angles found significant shape-specific information, in lateral occipital, posterior parietal and frontal areas. Thus, to test whether these areas retain the memorized shape in a rotation-invariant manner, we probed the recorded brain activity for information about the specific rotational angle of the memorized shape. To facilitate this analysis, we again grouped trials with similar rotations of a given shape into three bins, based on the assumption that similar rotations of the same shape are similar when represented retinotopically (see Methods; for details). We then tested whether the activity pattern within a given searchlight showed reliable differences across the different rotation bins for a given shape. Using the same thresholds as before, we did not find any area having significant levels of pattern distinctness D , as expected from an invariant signal. This held true even when multiple comparisons correction was restricted to only account for the voxels found in the first analysis (i.e. ‘small volume correction’). Thus, our results can be interpreted as indicative of rotation-invariant representations of memorized shapes in later occipital, posterior parietal and frontal areas when subjects perform a memory task where rotation is irrelevant.

Notably, the cvMANOVA-based rotation-specificity analysis using rotation-bins has an important caveat when compared to more conventional classifier-based ‘decoding’ analyses: It probes differences between different rotational views against the within-class (i.e. within-bin) variance and can therefore be assumed to be less powerful than a straight forward classification approach (which simply averages within-run variation). Thus, to confirm our results, we again used a time-resolved searchlight classification approach probing the acquired data for rotation-specific information (see Methods; for details). We found no evidence for such information in any of the areas reported previously, thus, confirming the evidence for a rotation-invariant code during the rotation-invariant task.

Shape representations during the rotation-specific (non-invariant) task

Then, we asked whether shapes were retained in a rotation-invariant format when retaining the rotation was necessary for the task. For this purpose, we analyzed data from the second group of subjects which performed a version of the task where the specific rotational angle of the memorized shape had to be remembered. This analysis was identical to the one used to identify invariant representations during the invariant task.

We found that even if subjects had to retain the rotation, shape-identity was retained in a rotation-invariant form as indicated by values of pattern distinctness D significantly larger than zero for the shape-identity analysis (one-sample t-test, $p < 0.001$, cluster-corrected $p_{(FWE)} < 0.05$, Figure 4B) and no significant rotation-specific information. The overall pattern of results during the rotation-specific task was largely identical to the results found during the rotation-invariant task. In particular, we found clusters which held significant information about the memorized shape in occipito-temporal, posterior parietal and frontal cortices, with peaks in the vicinity of LOC ($XYZ_{(MNI)}$; [-54 8 34] & [52 8 32], $T_{(21)} = 4.57$ & 5.71), the intraparietal sulcus ($XYZ_{(MNI)}$;

[-16 -74 52] & [34 -62 54], $T_{(21)} = 4.23$ & 4.3) and in right premotor cortex (XYZ_(MNI); [22 -4 66], $T_{(21)} = 5.48$). The left premotor cortex showed D values significantly larger than zero but the resulting cluster did not reach sufficient size to overcome cluster-correction (one-sample t-test, $p < 0.001$, XYZ_(MNI); [-20 -12 56], $T_{(21)} = 5.58$). All three areas also showed significantly larger D for memorized shapes as compared to not-memorized but shown shapes (paired t-test, LOC: $T_{(21)} = 3.05$ & 2.71 ; PPC: $T_{(21)} = 3.44$ & 2.98 ; FEF: $T_{(21)} = 3.11$ & 4.06 , $p < 0.05$, see Figure 5A) and significant across-rotation classification of shape-identity (LOC: $T_{(21)} = 3.56$ & 2.92 ; PPC: $T_{(21)} = 4.12$ & 4.44 ; FEF: $T_{(21)} = 4.02$ & 3.52 , $p < 0.05$, see Figure 5B). This effect, however, seems to be absent in left LOC when testing data from the rotation-specific task for peak voxels that hold information in the rotation-invariant task and vice versa in left and right LOC (see Figure 5). This could suggest that separate neural populations in LOC support memory function during the rotation-invariant and the rotation-specific task.

On the whole-brain level, a direct comparison (unpaired t-test) of the shape-specific and rotation-specific pattern distinctness values D between the rotation-invariant versus the rotation-specific task did not reveal any significant differences between the two groups of subjects ($p > 0.001$), suggesting that storage is largely unaffected by the task. Notably, posterior parietal cortex showed significantly larger values of D in the rotation-invariant task as compared to the rotation-specific task (unpaired t-test, $T = 3.11$, $p < 0.05$) when probing the areas reported above using a region-of-interest approach (see Figure 5). Decoding accuracy values derived from the cross-classification analyses showed no significant differences between the two tasks.

As for data from the rotation-invariant task, we also used a more conventional time-resolved searchlight-classification analysis to test the acquired data for rotation-specific information (see Methods; for details). When subjects had to specifically memorize the rotational

state of a given sample, we found no significant information of rotation specific images. Given previous reports of rotation-specific delay-period information (Harrison and Tong 2009; Serences et al. 2009; Christophel et al. 2015; Ester et al. 2015) we performed an exploratory analysis where we lowered the threshold and used an uncorrected threshold ($p < 0.001$). This identified a medial cluster in the precuneus ($XYZ_{(MNI)} [-2 -64 30]$, $T_{(20)} = 3.99$, $p < 0.001$, see Figure 6). When excluding the first time point which typically shows no information in multivariate analyses of fMRI data during working memory (Harrison and Tong 2009; Christophel et al. 2012) this cluster would be significant even after correction (cluster-corrected $p_{(FWE)} < 0.05$). This area did not overlap with any of the regions reported in rotation-invariant analyses. Furthermore, this region showed significantly higher decoding accuracies when decoding the rotational state of a given shape during the rotation-specific task as compared to the rotation-invariant task (two-sample t-test, $T_{(41)} = 3.55$, $p < 0.001$, see Figure 5). In contrast, the same area showed no significant cross-classification between shapes during the rotation-specific task. Decoding accuracy for such a rotation-invariant cross-classification was found to be significantly above chance during the rotation-specific task ($T_{(21)} = 3.17$, $p < 0.05$) and significantly larger than the rotation-invariant task (two-sample t-test, $T_{(41)} = 2.08$, $p < 0.05$). No other area showed significantly larger rotation-specific information in either of the tasks compared to the other task.

Frontal eye field localizer and region-of-interest analysis

The results in frontal cortices were of particular interest due to their implication for theories postulating storage in frontal cortex. Some authors postulate a unique role of dorsolateral prefrontal cortex (DLPFC) as a generalized working memory store (Goldman-Rakic 1995;

Petrides 1995; Funahashi 2006; Courtney et al. 2007). Others attribute a specific role in the storage of specific features to specific regions in prefrontal cortex as part of storage system distributed across the whole neocortex (e.g. Zimmer 2008).

Notably, the anatomical definition of the DLPFC with respect to Brodmann areas (Brodmann 1909) varies, with some authors referring to the lateral aspects of areas 9 and 46 (e.g. Hoshi 2006) and some also including area 8 (see Miller and Cohen 2001). In contrast, the results in the current study are almost exclusively localized in area 6, typically referred to as premotor cortex (Barbas and Pandya 1987; Wise 1985; see Campbell 1905). More importantly, the findings in premotor cortices overlap with the visual field maps constituting the frontal eye fields (Ferrier 1875) which are also situated in area 6 in humans (Petit et al. 1999) and sometimes are referred to as inferior and superior PCS (Jerde et al. 2012). In particular, the peaks found ($XYZ_{(MNI)}$; Group 1: [-54 8 34] & [52 8 32]; Group 2: [-20 -12 56] & [22 -4 66]) are within proximity of the results found during saccadic and pursuit eye movements (17 studies reviewed in Petit et al. 1999 $XYZ_{(MNI)}$; [-34 -14 50] & [32 -12 49]), sites where cortical stimulation elicits eye movements (Average electrode site in 2 patients, Blanke et al. 2000 $XYZ_{(MNI)}$; [-40 -9 54]) and regions of the frontal cortex involved in bottom-up saliency processing (Bogler et al. 2011; $XYZ_{(MNI)}$; [-27 -3 36] & [27 -9 51]).

To further explore this, subjects in the second group performed a short eye movement task. Instructed by abstract cues, subjects performed horizontal and vertical eye movements intermitted by periods of fixation. We estimated eye-movement related responses using hemodynamic-response function modeling and contrasted the corresponding parameter estimates against baseline on the group level (for details see Methods). Consistent with our main results, we found regions in premotor cortices with increased activation during eye movements

(XYZ_(MNI): [-56 4 44] & [54 -2 44]). We used these activation maps to generate two group-level left and right hemispheric regions-of-interest (ROIs) for the frontal eye fields (see Methods; for details). These ROIs showed large overlaps with content-specific activity patterns during the rotation-invariant and the rotation-specific task (Figure 7A, showing a collapsed analysis treating all subjects as one group, one-sample t-test, $p < 0.0001$, cluster-corrected $p_{(FWE)} < 0.05$).

To further investigate whether areas activated by eye movements were carrying memory-content-specific information, we used a pattern distinctness analysis which only considered data within these cortical ROIs. The estimated pattern distinctness values D within these ROIs were tested against chance (zero) on group level. We found invariant content-specific signals in FEF ROIs (one-sample t-test; Figure 7B) across both hemispheres in the rotation-invariant task ($T_{(21)} = 3.98$ & 3.47 , $p < 0.05$) and in right FEF for the rotation-specific task ($T_{(21)} = 4.48$, $p < 0.05$).

Thus, all evidence considered, we show that the content-specific activation patterns identified here reside in the frontal eye fields rather than the DLPFC. These results cannot be easily explained by stimulus-related eye movements, as the signals were found long after the presentation of the sample stimuli and before the onset of the target stimuli. Furthermore, stimulus-specific eye movements (i.e. tracing the outlines) would be specific to the rotational view shown and therefore would not lead to rotation-invariant signals.

Time-courses

For descriptive purposes, to explore the temporal evolution of content-specific signals during the delay period, we extracted the pattern distinctness D from the normalized and smoothed searchlight maps for each time-point separately. To provide an unbiased estimate of the effect size, we show data for regions-of-interest that were defined using independent data by

plotting data from the rotation-invariant task shown in regions defined using the rotation-specific task and vice versa. For *comparisons* between the two tasks, we use regions-of-interest defined using the collapsed data set. Please note that this approach resulted identical and near identical regions-of-interest for LOC and FEF in the analyses of the data for the rotation-invariant task and the collapsed data. For descriptive purposes only, we also report data from a given task for the regions defined within the same data. Data within regions defined using the same data cannot be used as an unbiased estimator of effect size.

These D-values are shown in Figure 7C separately for the lateral occipital regions, posterior parietal cortex and the frontal eye fields. Generally, all areas show little or no differences from zero during the initial seconds (most likely due to the hemodynamic delay), peak early and show a sustained response, similar to time courses reported previously (e.g. in Christophel et al. 2012; Christophel and Haynes 2014). The absolute values of D are slightly lower but of the same order of magnitude than those reported for category-specific information of perceived objects in the work by Allefeld and Colleagues (2014).

Correlations with individual differences in behavioral performance

Finally, we also asked whether the content-specific information retrievable from the acquired delay-period data covaries with behavioral performance. For this purpose, we correlated the individual behavioral accuracy (% correct) with pattern distinctness and decoding accuracy values computed using cvMANOVA (both shape-specific and rotation-specific), shape cross-classification and rotation classification analyses similar to prior work (Ester et al. 2013). Even though we computed these correlations for all areas and various content-specificity measures reported in this manuscript, we only found significant correlations of behavioral performance with shape-specific pattern distinctness values in the precuneus and only during the rotation-

invariant task (see Figure 5, $r = 0.54$; $p < 0.05$). It is important to mention, however, that performance correlations with content-specificity measures have so only found for relatively simple working memory tasks (e.g. Ester et al. 2013). In more complex tasks, like the one used here, where performance can depend on both the precision of memorization and the precision of the comparison of sample and test stimuli, such correlations have not been found (Christophel et al. 2012).

Discussion

Our results have implications for the question *where* and *how* information is held in working memory. Consistent with recent decoding studies of working memory (Harrison and Tong 2009; Serences et al. 2009; Christophel et al. 2012; Riggall and Postle 2012; Emrich et al. 2013; Lee et al. 2013; Christophel and Haynes 2014), we show that when artificial random shapes are memorized, lateral occipital cortex, posterior parietal cortices and the precuneus carry representations of these memorized contents. Importantly, we extend prior findings by showing that content-specific signals for visual working memory contents can also be found in a small region in (pre-) frontal cortex. We show, however, that this area overlaps with regions activated by a frontal eye field localizer task (Figure 7A & 7B) and thus presumably with areas where electrical stimulation elicit eye movements (Blanke et al. 2000). It can be concluded that the frontal eye fields (originally discovered by Ferrier 1875) hold shape-specific activity identifiable during working memory delays. Please note that the information we find in FEF is obtained during the delay period and thus is unlikely to reflect explicit eye-movements.

Thus, we demonstrate that the storage of visual stimuli held in working memory does indeed extend into frontal cortex as suggested by early work in non-human primates (Goldman-Rakic 1995; Petrides 1995). In contrast to this early work, our results are specifically localized to the frontal eye fields. Importantly, frontal eye field neurons do not only discharge during saccadic and pursuit eye movements (Bizzi 1968; Bizzi and Schiller 1970) but are also responsive to visual stimulation in the absence of eye movements (Bruce and Goldberg 1985). These visual responses form retinotopic maps of the visual field (Mohler et al. 1973) similar to maps found in the posterior parietal cortex. Our results suggest that such perception-driven neural populations within the (motor-driven) frontal eye fields can form lasting representations for working memory.

This is consistent with a recent study that found representations of memorized locations in the visual field putatively localized in the frontal eye fields (Jerde et al. 2012). Beyond this simple spatial signal, evidence for content-specific memory signals in human frontal cortices was – until recently – limited to signals of a clearly non-sensory nature (Spitzer and Blankenburg 2012; Lee et al. 2013; Spitzer et al. 2014). More recently, Ester and colleagues (2015) showed that frontal cortices carry information about memorized gratings, most reliably in the vicinity of the frontal eye fields, albeit leaving open whether the representation of gratings in these areas uses a spatial or a low-level visual code. Extending these findings our results show that sensory driven populations in the frontal eye fields are not only capable of representing spatial position but can also code for visual shapes independently of spatial position in an invariant manner. In non-human primates, the frontal eye fields have also been shown to code for spatial location during memory (Armstrong et al. 2009) while representations of simple canonical shapes haven only been found in neurons recorded from adjacent principal sulcus regions (Meyer et al. 2011;

Meyers et al. 2012). Information about complex shapes were only found in inferotemporal areas (Miyashita and Chang 1988). We extend this prior work by suggesting that visual working memory content-specificity is primarily localized to the retinotopic frontal eye fields rather than the dorsolateral prefrontal cortex that is considered to be less visually responsive.

Regarding *how* these stimuli are retained, we found that representations held in working memory generalize across rotations in posterior parietal cortex, lateral occipital areas and the frontal eye fields. To demonstrate this generalization, we used the recently developed cvMANOVA approach (Allefeld and Haynes 2014) estimating pattern distinctness to show that delay-period activation patterns carry information about the memorized shape without representing its rotational view. Notably, we did not test whether the same representations also generalize across variations of other properties of the shapes like size or location, leaving this as an open question.

Importantly, rotation-invariant coding of working memory contents was prevalent in lateral occipital, posterior parietal and frontal regions even when subjects had to memorize the rotational state to perform (in the rotation-specific task). This evidence stands in contrast to prior work which showed that, during working memory, posterior parietal cortex and the frontal eye fields carry rotation information (Christophel et al. 2015; Ester et al. 2015). Two explanations can be considered to interpret this discrepancy: On the one hand, this contradicting evidence could be a result a difference in experimental tasks and designs. As a major difference, prior work (Christophel et al. 2015; Ester et al. 2015) used a limited set of rotations (four rotations in Christophel et al. 2015; nine rotations with a $\pm 1-5^\circ$ jitter in Ester et al. 2015) whereas the current study employed random rotations. Even though we intended to discourage the use of invariant strategies by using random rotations (see above), it is possible that this choice in design had the

opposite effect. Notably, rotation-specific motion information during working memory in sensory areas was also found when a few fixed rotations were interleaved with many random rotations (Emrich et al. 2013).

On the other hand, the difference between the current and prior work could be explained by the type of stimulus used. Prior work (Christophel et al. 2015; Ester et al. 2015) did not employ stimuli that can be considered as shapes (or even objects) but employed abstract grating or pattern stimuli. Our finding could suggest that the encoding processes underlying visual working memory storage of shapes and objects (but not gratings and patterns) in these areas neglect the original viewing conditions and directly generate representations based on an invariant code. In day-to-day life, memorized objects oftentimes change their location relative to the observer before recall rendering the exact viewing conditions irrelevant. In such a situation, memory contents stored in a strictly retinotopic format would be impractical. They would necessitate some form of mental rotation (Shepard and Metzler 1971) of the memorized contents to allow for comparison. Instead, the invariant encoding of visual shapes in working memory could be achieved by using an object-centered coordinate systems (Marr and Nishihara 1978) or by using canonical orientations (Tarr and Pinker 1989) to allow for direct comparisons. Our evidence suggests that such a process would be limited to stimuli that have a clear unifying outline or are perceived as *one* persistent object for other reasons.

During perception, invariant representations in ventral stream areas (like the LOC) are believed to be based on an element-by-element code, generating larger shapes from smaller fragments in 2D (Brincat and Connor 2004) or 3D (Yamane et al. 2008). Such a format, with shapes sharing similar elements forming a columnar-level shape space (Fujita et al. 1992; Tanaka 1993) could also be utilized during working memory in this study. In contrast, both intraparietal

areas and the frontal eye fields show a retinotopic organization during perception (Tootell et al. 1998; Sereno et al. 2001; Hagler Jr et al. 2006; Wandell et al. 2007). During working memory, instead of coding for the retained shape as it was shown, these areas might retain the shape of visual objects in a rectified or canonical (Tarr and Pinker 1989; see above) rotational state using these retinotopic maps as ‘canvas’. Furthermore, the shapes might not be represented as-a-whole in these retinotopic areas but only by coding the relative positions of salient parts of the stimulus (see Bogler et al. 2011; Christophel et al. 2012; Katshu and d’Avossa 2014) while the details of the shape are retained within LOC. Notably, invariant representations of visual objects in posterior parietal cortex have also been reported during perception (Konen and Kastner 2008), questioning whether this mechanism of representation is unique to working memory.

Recent behavioral work suggests the coexistence of both viewpoint-specific and viewpoint-invariant stores for visual working memory (Wood 2009, 2011). This idea is further supported by our rotation-specific analyses: When subjects were performing the rotation-invariant task and the rotational view of the sample during presentation was irrelevant, we found no evidence for no such rotation-specific information. Only if the retention of the viewing conditions were strictly necessary to perform in the task (in the second group), we found some limited evidence for the retention of these viewing conditions in neural circuits. These results were localized to the precuneus, which has been postulated to carry viewpoint-specific information before during imagery (Fletcher et al. 1995; Burgess 2002), but not in lateral occipital regions, posterior parietal cortex and the frontal eye fields. During the rotation-invariant task, however, the same region showed evidence for rotation-invariant storage that was indistinguishable from the results in other areas. Thus, rotation-invariant and rotation-specific information were localized to different cortical regions, supporting the idea of separate stores (Wood 2009, 2011), but the

precuneus was capable of retaining either rotation-specific or rotation-invariant representations, depending on current task demands.

A different area that could potentially serve as a view-specific store is the primary visual cortex (V1). Several multivariate decoding studies of working memory report significant cross-classification between working memory and perceptual signals (Harrison and Tong 2009; Riggall and Postle 2012; Albers et al. 2013; Emrich et al. 2013), suggesting that representations during visual working memory rely on a neural code that is similar or identical to that underlying visual perception. Furthermore, a recent study showed that the orientation-specific responses during working memory for grating stimuli follow tuning curves similar to responses during perception (Ester et al. 2013).

In a noteworthy study, however, Ester and Colleagues (2009; see also Jehee et al. 2011; Serences and Boynton 2007) showed that it is possible to decode the orientation of a memorized grating originally shown either to the left or right visual field from ipsilateral areas in visual cortex coding for the opposite side. This finding speaks against a strictly retinotopic representation of memorized stimuli even in early visual cortex, at least when - like in the study by Ester and Colleagues (2009) - their spatial position was irrelevant for the task. A more recent study (Pratte and Tong 2014), however, found that the spread of orientation-specific information was limited to extrastriate visual areas V3-V4 when subjects had to memorize two orientations shown in the left and right visual fields. Furthermore, the spreading of information across the cortical hemispheres of V1 proposed by the Ester and colleagues (2009) might not allow for generalized representations of spatially inhomogeneous stimuli like the ones used in the current study. Therefore, the cross-hemisphere spread of information in V1 might be limited to local features like gratings, texture and color (but see Williams et al. 2008).

Here, we neither found rotation-invariant nor rotation-specific information in V1, suggesting either that neural populations in V1 do not contribute to the retention of complex shapes and their rotational view or that the specific tasks and design prevented us from finding potential content-specific signals in V1. In particular, each trial included a not-memorized control shape shown at an identical position of the visual field in addition to a star-shaped mask. These additional stimuli could have suppressed or concealed rotation-specific activity patterns in V1, which might otherwise enable subjects to perform in the rotation-specific task. Furthermore, rotation-specific information in memory has been reported previously in early visual areas using a small set of possible rotational states (Albers et al. 2013; Christophel et al. 2015). The current study was primarily designed to investigate invariant rather than rotation-specific responses and used random rotations in each trial to discourage simplified memorization strategies. This made the use of simple contrasts between different rotations impossible. Thus, whether rotation-specific signals found previously can be explained by the categorical nature of rotation in prior studies or whether rotation-specific information is rendered indiscernible in the current study by using random rotations is a question of further research.

Conclusions

Our conclusions are threefold: First, our findings represent evidence for visual working memory stores forming a distributed network of brain regions including occipital, parietal and frontal cortices. Second, the visual shape representations found here in frontal regions are specifically localized in the frontal eye fields, which have so far only been implied in memory

for spatial position and orientation. Third, visual working memory storage in these areas might be based on an invariant coding scheme.

Figure Legends

Figure 1: Random Shape Memory Task. (A) The timeline (in seconds) of an experimental trial is shown horizontally. Stimulus displays are shown with onsets indicated by vertical lines and duration of stimulus displays are given as horizontal lines. In each trial, subjects were cued to memorize one out of two decagons shown (drawn from a set of six stimuli). The two shapes were shown sequentially followed by a *retro-cue* ('1' or '2') indicating which of the two should be memorized. These stimuli were generated randomly based on an algorithm that allowed to parametrically vary shape similarity (for details see Figure 2). The cue was surrounded by a star-shaped mask introduced to suppress activity merely related to stimulus presentation. The to-be-memorized (cued) shape was shown in a random in-plane rotation, whereas the not memorized shape was always shown in the same view. After an extended *delay* (10 s), subjects performed a difficult similarity judgment. Two test stimuli were shown left and right of fixation. Subjects were asked to identify which of the two stimuli was more similar to the memorized shape. (B) A unique set of six memory shapes for each subject was shown (one example set is shown, see *design*) and rotational angle was randomized in each trial (shown shaded and overlapping to emphasize the random nature). Each subject performed four runs of 60 trials each (10 per memorized sample) in the scanner. There were two groups of subjects either performing only a rotation-invariant or only the rotation-specific version of this similarity judgment task. This was done to avoid carryover of strategies between the two tasks. (C) For the *rotation-invariant task* the rotation of the target stimulus was presented at a random rotational angle. The foil stimulus was uncorrelated with the target at all rotational angles. Examples for target and foil stimuli are shown. (D) In the *rotation-specific task*, subjects had to memorize the shape of the cued sample in the specific rotation shown. Here, the target stimulus was always shown in the same rotational angle as the memorized shape. In order to explicitly discourage subjects from using a rotation-invariant strategy, it was possible for the foil to have a higher similarity to the memorized shape when appropriately rotated, so the correct answer could only be obtained by taking the original rotational angle into account. In addition, one out of four trials was a catch trial where the foil was identical to the memorized stimulus but shown in a random rotation. Examples for target and foil stimuli are shown. In both tasks, the target stimulus rotation was chosen randomly.

Figure 2: Stimulus generation and similarity metrics. (A) To randomly generate ten-sided shapes (decagons), we assigned one out of 10 evenly-spaced numbers (from 0.28 to 1) randomly to each corner. These values were used to determine the distance from each corner to the center. All points were 36° radial angle apart from each other relative to the center. The shapes were then recentered and resized to control for slight variations across different stimuli in center-of-mass and surface area. Shapes were shown in white on a black circular patch. (B) For the subsequent task, we estimated shape similarity using two different measures. Rotation-specific similarity was defined by correlating the corner-to-center distance with each corner aligned in the given view. Rotation-invariant correlation was computed by shifting the order of corner-to-center distance, and using the maximal correlation and thereby implicitly computing the similarity for the shapes when rotated optimally.

Figure 3: Behavioral Results. Percentages of correct responses and reaction times (in seconds) are shown for each individual subject in the rotation-invariant (grey, left) and the rotation-specific (black, right) task. Filled squares indicate the mean per subject and empty circles show data from each of the four runs.

Figure 4: Regions of the human brain carrying rotation-invariant information about the memorized shape during the delay period. (A) Lateral views of a semi-transparent rendered surface of the human brain showing in green regions which showed significant (one-sample t-test, $p < 0.001$, cluster-corrected $p_{(FWE)} < 0.05$) statistically significant information (pattern distinctness D) about the memorized shape collapsing across different rotational views during the rotation-invariant task. (B) Areas shown in orange for an identical analysis performed on data for the rotation-specific task. None of these areas showed significant information about the rotational state of the shape in either of the tasks. No additional clusters are visible when the renders are shown from a different perspective.

Figure 5: Effect sizes and between-subject variation (SEM) reported for shape-specific (rotation-invariant) cvMANOVA (A) and shape-specific (rotation-invariant) and rotation-specific classification (B) analyses. Data shown mainly for illustrative purposes is extracted from pattern distinctness or classification accuracy maps generated by a searchlight-based algorithm. The left panel (A) shows pairs of bars where the left bar indicates shape-specific (rotation-invariant) information for the remembered item and the right bar represents shape-specific information for the not-remembered item. In the right panel (B), pairs of bars show shape-specific cross-classification accuracy for classifiers trained and tested on different rotations of the same shape on the left and classification accuracy for a classifier that distinguishes between these different rotations of the same shape on the right. The individual plots represent data for occipito-temporal (top), parietal (middle) and frontal (bottom) voxels. Voxels from which data is shown are defined from prior analyses (green, white, orange) or prior work and anatomy (gray). To provide an unbiased estimate of the effect size, we show data for regions-of-interest that were defined using independent data by plotting data from the rotation-invariant task shown in regions defined using the rotation-specific task and vice versa. For comparisons between the two tasks, we use regions-of-interest defined using the collapsed data set. For descriptive purposes only, we also report data from a given task for the regions defined within the same data for the sake of completeness. Data within regions defined using the same data cannot be used as an unbiased estimator of effect size and the statistics shown are redundant to the analyses defining them. Error bars indicate SEM. Asterisks atop of bars indicate significant above chance information (one-sample t-test, $p < 0.05$), asterisks atop of solid brackets indicate significant differences in pattern distinctness D for memorized vs. not-memorized items for the same voxels (paired t-test, $p < 0.05$), asterisks over dashed brackets show significant between-group comparisons for the same voxels between subjects doing the rotation-specific and the rotation-invariant task (unpaired t-test, $p < 0.05$). Plus signs over bars and linked to them via a dotted line indicate significant correlation of content-specific information and behavioral accuracy ($p < 0.05$). XYZ(MNI) for occipito-temporal (left to right, LOC: [-44 -70 -4], [-44 -70 -4], [-28 -88 16], [46 -82 10], [46 -82 10] & [42 -84 -2]; calcarine sulcus: [-1 -82 10] & [10 -88 10]), parietal (left to right, PPC: [-22, -72, 50], [-24, -72, 44], [-16, -74, 52], [34, -50,

48], [20, -72, 44] & [34 -62 54]; Precuneus: [-2, -64, 30] & [2, -64, 30]) and frontal (FEF: [-54 8 34], [-52 6 32], [-20 -12 56], [52 8 32], [52 8 40] & [22 -4 66]; DLPFC: [-44 52 21] & [44 52 21]) regions refer to the voxels from which data is extracted from the searchlight maps. Please note that in LOC data for two voxels are plotted twice as they were identified as peaks in analyses of the rotation-invariant task data and of all data collapsed.

Figure 6. Regions of the brain showing information about the rotational view of the memorized shape during presentation in searchlight classification analysis using data from the rotation-specific task. Medial views of a semi-transparent rendered surface of the human brain showing in orange regions which showed above chance classification for support vector machine classifiers trained to distinguish between different views of the same shape and tested on independent data. The same analysis performed on data from the rotation-invariant task yielded no results at the same threshold (see also Figure 5). No additional clusters are visible when the renders are shown from a different perspective.

Figure 7: Frontal eye field region-of-interest analyses and time-series of invariant content-specific information across all areas. (A) Lateral views of a rendered representation of the human brain showing in yellow the results of the pattern distinctness analysis probing the recorded fMRI data for rotation invariant representations collapsed across the rotation-invariant and the rotation-specific task (one-sample t-test, $p < 0.0001$, cluster-corrected $p_{(FWE)} < 0.05$). In red we show two overlapping left and right hemispheric frontal eye field regions-of-interests (ROIs) which were defined on the group level using an eye movement localizer task (see methods section). (B) ROI pattern distinctness probing the frontal eye field for rotation-invariant information about the memorized shape in the two tasks. Bar plots show pattern distinctness D . Asterisk indicate significance (one-sample t-test, $p < 0.05$). (C) Delay-period time-series of pattern distinctness D across the three areas found in the main analyses (Figure 4) shown to illustrate the temporal evolution of content-specific activity. D -values are extracted from normalized and smoothed searchlight maps of D and are shown for statistical peaks reported during the rotation-invariant task (green), the rotation-specific task (orange) and a version of the analysis which collapsed across the two tasks (empty squares connected by lines). To provide an unbiased estimate of the effect size, we show data for regions-of-interest that were defined using independent data by plotting data from the rotation-invariant task shown in regions defined using the rotation-specific task and vice versa. For *comparisons* between the two tasks, we use regions-of-interest defined using the collapsed data set. For descriptive purposes only, we also report data from a given task for the regions defined within the same data for the sake of completeness. Data within regions defined using the same data cannot be used as an unbiased estimator of effect size.

References

- Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP. 2013. Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Curr Biol.* 23:1427–1431.
- Allefeld C, Haynes J-D. 2014. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *NeuroImage.* 89:345–357.
- Armstrong KM, Chang MH, Moore T. 2009. Selection and Maintenance of Spatial Information by Frontal Eye Field Neurons. *J Neurosci.* 29:15621–15629.
- Ashburner J, Friston KJ. 2005. Unified segmentation. *Neuroimage.* 26:839–851.
- Baddeley A. 1992. Working memory. *Science.* 255:556–559.
- Baddeley AD. 1986. Working Memory. Oxford (UK): Clarendon Press.
- Barbas H, Pandya DN. 1987. Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *J Comp Neurol.* 256:211–228.
- Bettencourt KC, Xu Y. 2016. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat Neurosci.* 19:150–157.
- Bizzi DE. 1968. Discharge of frontal eye field neurons during saccadic and following eye movements in unanesthetized monkeys. *Exp Brain Res.* 6:69–80.
- Bizzi E, Schiller PH. 1970. Single unit activity in the frontal eye fields of unanesthetized monkeys during eye and head movement. *Exp Brain Res.* 10:151–158.
- Blanke O, Spinelli L, Thut G, Michel CM, Perrig S, Landis T, Seeck M. 2000. Location of the human frontal eye field as defined by electrical cortical stimulation: anatomical, functional and electrophysiological characteristics. *Neuroreport.* 11:1907–1913.
- Bogler C, Bode S, Haynes J-D. 2011. Decoding Successive Computational Stages of Saliency Processing. *Curr Biol.* 21:1667–1671.

Brincat SL, Connor CE. 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci.* 7:880–886.

Brodmann K. 1909. *Vergleichende Lokalisationslehre der Großhirnrinde.* Springer.

Bruce CJ, Goldberg ME. 1985. Primate frontal eye fields. I. Single neurons discharging before saccades. *J Neurophysiol.* 53:603–635.

Burgess N. 2002. The hippocampus, space, and viewpoints in episodic memory. *Q J Exp Psychol A.* 55:1057–1080.

Campbell AW. 1905. *Histological studies on the localisation of cerebral function.* Cambridge, University Press.

Christophel TB, Cichy RM, Hebart MN, Haynes J-D. 2015. Parietal and early visual cortices encode working memory content across mental transformations. *Neuroimage.* 106:198–206.

Christophel TB, Haynes J-D. 2014. Decoding complex flow-field patterns in visual working memory. *Neuroimage.* 91:43–51.

Christophel TB, Hebart MN, Haynes J-D. 2012. Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. *J Neurosci.* 32:12983–12989.

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J-D. 2017. The Distributed Nature of Working Memory. *Trends in Cognitive Sciences.* 0.

Cichy RM, Chen Y, Haynes JD. 2011. Encoding the identity and location of objects in human LOC. *Neuroimage.* 54:2297–2307.

Courtney SM, Roth JK, Sala JB. 2007. A hierarchical biased-competition model of domain-dependent working memory maintenance and executive control. In: N. Osaka, Robert H.

Logie, D'Esposito M, editors. *The Cognitive Neuroscience of Working Memory*. Oxford: Oxford Univ. Pres. p. 369–384.

Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*. 19:261–270.

Emrich SM, Riggall AC, LaRocque JJ, Postle BR. 2013. Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. *J Neurosci*. 33:6516–6523.

Ester EF, Anderson DE, Serences JT, Awh E. 2013. A Neural Measure of Precision in Visual Working Memory. *J Cogn Neurosci*. 25:754–761.

Ester EF, Serences JT, Awh E. 2009. Spatially global representations in human primary visual cortex during working memory maintenance. *J Neurosci*. 29:15258–15265.

Ester EF, Sprague TC, Serences JT. 2015. Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*. 87:893–905.

Ferrier D. 1875. The Croonian Lecture - Experiments on the Brain of Monkeys. *Philos T R Soc Lond*. 165:433–488.

Fletcher PC, Frith CD, Baker SC, Shallice T, Frackowiak RSJ, Dolan RJ. 1995. The mind's eye—precuneus activation in memory-related imagery. *Neuroimage*. 2:195–200.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. 1991. Comparing Functional (PET) Images: The Assessment of Significant Change. *J Cereb Blood Flow Metab*. 11:690–699.

Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RS. 1994. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Map*. 2:189–210.

Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. 1994. Assessing the significance of focal activations using their spatial extent. *Hum Brain Map.* 1:210–220.

Fujita I, Tanaka K, Ito M, Cheng K. 1992. Columns for visual features of objects in monkey inferotemporal cortex. *Nature.* 360:343–346.

Funahashi S. 2006. Prefrontal cortex and working memory processes. *Neuroscience.* 139:251–261.

Goldman-Rakic PS. 1995. Architecture of the Prefrontal Cortex and the Central Executive. *Ann NY Acad Sci.* 769:71–84.

Grill-Spector K, Kourtzi Z, Kanwisher N. 2001. The lateral occipital complex and its role in object recognition. *Vision Res.* 41:1409–1422.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R. 1999. Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex. *Neuron.* 24:187–203.

Hagler Jr DJ, Saygin AP, Sereno MI. 2006. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage.* 33:1093–1103.

Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. *Nature.* 458:632–635.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science.* 293:2425–2430.

Haynes JD, Rees G. 2005a. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci.* 8:686–691.

Haynes J-D, Rees G. 2005b. Predicting the Stream of Consciousness from Activity in Human Visual Cortex. *Curr Biol.* 15:1301–1307.

Haynes J-D, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat Rev Neurosci.* 7:523–534.

Hebart MN, Görgen K, Haynes J-D. 2015. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front Neuroinform.* 8.

Hoshi E. 2006. Functional specialization within the dorsolateral prefrontal cortex: A review of anatomical and physiological studies of non-human primates. *Neurosci Res.* 54:73–84.

Hyun J-S, Luck SJ. 2007. Visual working memory as the substrate for mental rotation. *Psychon Bull Rev.* 14:154–158.

Jacob SN, Nieder A. 2014. Complementary Roles for Primate Frontal and Parietal Cortex in Guarding Working Memory from Distractor Stimuli. *Neuron.* 83:226–237.

Jehee JFM, Brady DK, Tong F. 2011. Attention Improves Encoding of Task-Relevant Features in the Human Visual Cortex. *J Neurosci.* 31:8210–8219.

Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE. 2012. Prioritized Maps of Space in Human Frontoparietal Cortex. *J Neurosci.* 32:17382–17390.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci.* 8:679–685.

Kanwisher N, Woods RP, Iacoboni M, Mazziotta JC. 1997. A Locus in Human Extrastriate Cortex for Visual Shape Analysis. *J Cogn Neurosci.* 9:133–142.

Katshu MZUH, d’Avossa G. 2014. Fine-Grained, Local Maps and Coarse, Global Representations Support Human Spatial Working Memory. *PLoS ONE.* 9:e107969.

Konen CS, Kastner S. 2008. Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci.* 11:224–231.

Kourtzi Z, Erb M, Grodd W, Bühlhoff HH. 2003. Representation of the Perceived 3-D Object Shape in the Human Lateral Occipital Complex. *Cereb Cortex.* 13:911–920.

Kourtzi Z, Kanwisher N. 2001. Representation of Perceived Object Shape by the Human Lateral Occipital Complex. *Science.* 293:1506–1509.

Kriegeskorte N, Cusack R, Bandettini P. 2010. How does an fMRI voxel sample the neuronal activity pattern: Compact-kernel or complex spatiotemporal filter? *NeuroImage.* 49:1965–1976.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA.* 103:3863–3868.

Lara AH, Wallis JD. 2014. Executive control processes underlying multi-item working memory. *Nat Neurosci.* 17:876–883.

Lee S-H, Kravitz DJ, Baker CI. 2013. Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci.* 16:997–999.

Logie RH. 2003. Spatial and visual working memory: A mental workspace. In: Irwin DE,, Ross BH, editors. *Cognitive vision. The psychology of learning and motivation.* San Diego, CA, US: Academic Press. p. 37–78.

Luck SJ. 2008. Visual short-term memory. In: Luck SJ,, Hollingsworth A, editors. *Visual memory.* Oxford (UK): Oxford University Press. p. 43–85.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci USA.* 92:8135–8139.

Marr D, Nishihara HK. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *P Roy Soc Lond B Bio.* 200:269–294.

Matsushima A, Tanaka M. 2014. Different Neuronal Computations of Spatial Working Memory for Multiple Locations within versus across Visual Hemifields. *J Neurosci.* 34:5621–5626.

Mendoza-Halliday D, Torres S, Martinez-Trujillo JC. 2014. Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci.* 17:1255–1262.

Meyer T, Qi X-L, Stanford TR, Constantinidis C. 2011. Stimulus Selectivity in Dorsal and Ventral Prefrontal Cortex after Training in Working Memory Tasks. *J Neurosci.* 31:6266–6276.

Meyers EM, Qi X-L, Constantinidis C. 2012. Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proc Natl Acad Sci USA.* 109:4651–4656.

Miller EK, Cohen JD. 2001. An Integrative Theory of Prefrontal Cortex Function. *Annu Rev Neurosci.* 24:167–202.

Miyashita Y, Chang HS. 1988. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature.* 331:68–70.

Mohler CW, Goldberg ME, Wurtz RH. 1973. Visual receptive fields of frontal eye field neurons. *Brain Res.* 61:385–389.

Nichols T, Hayasaka S. 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res.* 12:419–446.

Petit L, Dubois S, Tzourio N, Dejudin S, Crivello F, Michel C, Etard O, Denise P, Roucoux and A, Mazoyer B. 1999. PET study of the human foveal fixation system. *Hum Brain Map.* 8:28–43.

Petrides M. 1995. Functional Organization of the Human Frontal Cortex for Mnemonic Processing. *Ann NY Acad Sci.* 769:85–96.

Pratte MS, Tong F. 2014. Spatial specificity of working memory representations in the early visual cortex. *J Vis.* 14:22.

Riggall AC, Postle BR. 2012. The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging. *J Neurosci.* 32:12990–12998.

Serences JT, Boynton GM. 2007. Feature-Based Attentional Modulations in the Absence of Direct Visual Stimulation. *Neuron.* 55:301–312.

Serences JT, Ester EF, Vogel EK, Awh E. 2009. Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci.* 20:207–214.

Sereno MI, Pitzalis S, Martinez A. 2001. Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science.* 294:1350–1354.

Shepard RN, Metzler J. 1971. Mental Rotation of Three-Dimensional Objects. *Science.* 171:701–703.

Soon CS, Brass M, Heinze HJ, Haynes JD. 2008. Unconscious determinants of free decisions in the human brain. *Nat Neurosci.* 11:543–545.

Sperling G. 1960. The information available in brief visual presentations. *Psychol Monogr.* 74:1–29.

Spitzer B, Blankenburg F. 2012. Supramodal Parametric Working Memory Processing in Humans. *J Neurosci.* 32:3287–3295.

Spitzer B, Gloel M, Schmidt TT, Blankenburg F. 2014. Working memory coding of analog stimulus properties in the human prefrontal cortex. *Cereb Cortex.* 24:2229–2236.

- Sterzer P, Haynes J-D, Rees G. 2008. Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *J Vis.* 8:10.1-12.
- Tanaka K. 1993. Neuronal mechanisms of object recognition. *Science.* 262:685–688.
- Tarr MJ, Pinker S. 1989. Mental rotation and orientation-dependence in shape recognition. *Cogn Psychol.* 21:233–282.
- Timm NH. 2002. *Applied multivariate analysis.* New York (NY): Springer.
- Tootell RB., Hadjikhani N, Hall EK, Marrett S, Vanduffel W, Vaughan JT, Dale AM. 1998. The retinotopy of visual spatial attention. *Neuron.* 21:1409–1422.
- Wandell BA, Dumoulin SO, Brewer AA. 2007. Visual field maps in human cortex. *Neuron.* 56:366–383.
- Williams MA, Baker CI, Beeck HPO de, Shim WM, Dang S, Triantafyllou C, Kanwisher N. 2008. Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci.* 11:1439–1445.
- Wise SP. 1985. The Primate Premotor Cortex: Past, Present, and Preparatory. *Annu Rev Neurosci.* 8:1–19.
- Wood JN. 2009. Distinct Visual Working Memory Systems for View-Dependent and View-Invariant Representation. *PLoS ONE.* 4:e6601.
- Wood JN. 2011. A core knowledge architecture of visual working memory. *J Exp Psychol Human.* 37:357–381.
- Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE. 2008. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci.* 11:1352–1360.
- Zimmer HD. 2008. Visual and spatial working memory: From boxes to networks. *Neurosci Biobehav Rev.* 32:1373–1395.

